



Paper Type: Original Article

Beyond Shared Perception: A Comprehensive Review of Cooperative Perception Systems for Autonomous Driving

FirstName, LastName^{1,*}, FirstName, LastName

¹ Affiliation; Email Address;

² Affiliation; Email Address;

Citation:

Received: ----

Revised: ----

Accepted:---

LastName, Initial First Name., & LastName, Initial First Name. (Date).

Beyond Shared Perception: A Comprehensive Review of Cooperative

Perception Systems for Autonomous Driving. , Volume (Issue), PP.

Abstract

Cooperative Perception (CP) enables autonomous vehicles to share sensing information with other vehicles and infrastructure. This extends situational awareness beyond the range and line-of-sight limitations of onboard sensors, improving the safety and efficiency of Intelligent Transportation Systems (ITS). This paper provides a comprehensive review of CP. It covers multimodal sensing modalities, core system architectures, and Vehicle-to-Everything (V2X) communication technologies that enable perception data to be shared and coordinated across agents. It also reviews the data fusion strategies used to combine this shared information across the cooperative perception pipeline. The choice of fusion stage directly shapes the trade-off between detection accuracy, communication cost, and robustness under real-time, non-ideal network conditions, making it a central design decision for any CP system. We further compare widely used simulation and real-world CP datasets. Finally, we discuss open challenges in communication constraints, sensor heterogeneity, and security and privacy, as well as the transition from purely observational systems toward more coordinated, action-aware cooperative driving.

Keywords: Collective Perception Systems, V2X, Safety, Intelligent Transportation Systems, Autonomous vehicles.

1 | Introduction

Traffic safety is a fundamental requirement for human dignity, social stability, and sustainable development. Within the framework of Intelligent Transportation Systems (ITS), safety remains a persistent challenge as increasing automation introduces new vulnerabilities in sensor suites and communication networks. Adopting the "Vision Zero" paradigm, contemporary traffic safety is understood as a shared responsibility, necessitating

 Corresponding Author:



Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

a systemic design that integrates vehicles, infrastructure, and human agents in a collaborative safety environment [1, 2]. To realize these objectives, autonomous driving must transition from reactive, post hoc mitigation to proactive prevention. This evolution demands moving beyond simple reactive pipelines toward sophisticated architectures capable of reasoning under uncertainty. Proactive safety requires that predictive insights arrive with sufficient lead time to influence control decisions, allowing hazards to be anticipated and avoided before they materialize in the physical environment [2].

Conventional perception-centric systems have reached a plateau defined by their inherent physical and informational boundaries. Onboard sensors—including high-resolution cameras, LiDAR, and radar—are fundamentally restricted by line-of-sight (LoS) observations. In complex urban environments, the perceptual field of view is often compromised by occlusions caused by roadside structures, intricate urban layouts, and large vehicles [3, 4]. These spatial restrictions significantly shorten the ego vehicle's anticipation horizon. Furthermore, single-vehicle systems often struggle with informational boundaries: surrounding agents are inherently difficult to predict, their intent is frequently ambiguous, and sudden maneuvers may occur too close for safe intervention based on local data alone [5, 6].

In contrast, the foundational objective of Cooperative Perception (CP) is to merge distributed observations from multiple independent agents into a unified, collective representation of the environment [7–10]. This paradigm shifts the focus from isolated ego-vehicle sensing to a networked ecosystem, extending situational awareness beyond individual line-of-sight and sensor limits. Vehicle-to-Everything (V2X) technology, encompassing vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication, is the primary enabler of this ecosystem. By exchanging raw states, sensor features, and structured safety evidence, connected agents create a "collective intelligence" that facilitates proactive and coordinated maneuvers. A primary technical benefit of V2X is the extension of the perceptual horizon Beyond the Line Of Sight (BLOS). This spatial-temporal awareness allows vehicles to anticipate hidden objects or maneuvers in adjacent lanes—such as braking events occurring several vehicles ahead—long before they become visible to onboard sensors.

The realization of this cooperative architecture depends on three tightly coupled functional stages:

- I. Data collection: the continuous, multimodal acquisition of environmental data from onboard vehicle sensors—cameras, LiDAR, and radar—and from infrastructure-based sources such as Fixed Sensor Nodes (FSNs) and Roadside Units (RSUs). These infrastructure sensors are positioned at elevated vantage points to monitor high-risk zones, including complex intersections and areas with dense pedestrian activity [11, 12]
- II. Perception and fusion: the heterogeneous data streams are processed through dedicated pipelines for object detection, classification, and tracking. The resulting outputs are then fused into a unified scene representation, constructing a reliable, shared understanding of all traffic participants [13–15].
- III. Information distribution: fused perception outcomes are broadcast across the full V2X ecosystem, reaching connected vehicles, Vulnerable Road Users (VRUs), and infrastructure nodes such as traffic signals. This ensures that every agent acts upon consistent, safety-critical information [16].

Taken together, these three stages provide full-stack enhancements to autonomous driving architectures. The shared features detect objects far beyond immediate sensing ranges. For localization, data from neighboring agents and infrastructure supplements GPS signals, which is vital in urban canyons with poor satellite coverage. These enhancements propagate to the planning and control layers, where real-time awareness of neighboring trajectories and distributed AI inference enable adaptive decisions and precise trajectory execution [17–19].

This paper provides a comprehensive analysis of the CP ecosystem, synthesizing the technical requirements and architectural paradigms necessary to transition from ego-centric sensing to collective intelligence. The remainder of this paper is organized as follows. Section 2 explores the multimodal sensing layer, detailing visual, geometric, and spatiotemporal referencing techniques. Section 3 analyzes the structural organization of CP frameworks, ranging from cloud-centric centralization to decentralized V2V paradigms. Section 4 presents a deep dive into fusion paradigms, focusing on the trade-offs between raw data alignment and

semantic feature sharing. Section 5 discusses V2X communication technologies and the standardized message formats that enable cooperative awareness. Section 6 reviews the datasets and simulation tools currently driving research in the field. Section 7 identifies current challenges and future research gaps, emphasizing the shift from perception to coordinated action. Section 8 concludes the paper by summarizing the key findings and providing final remarks.

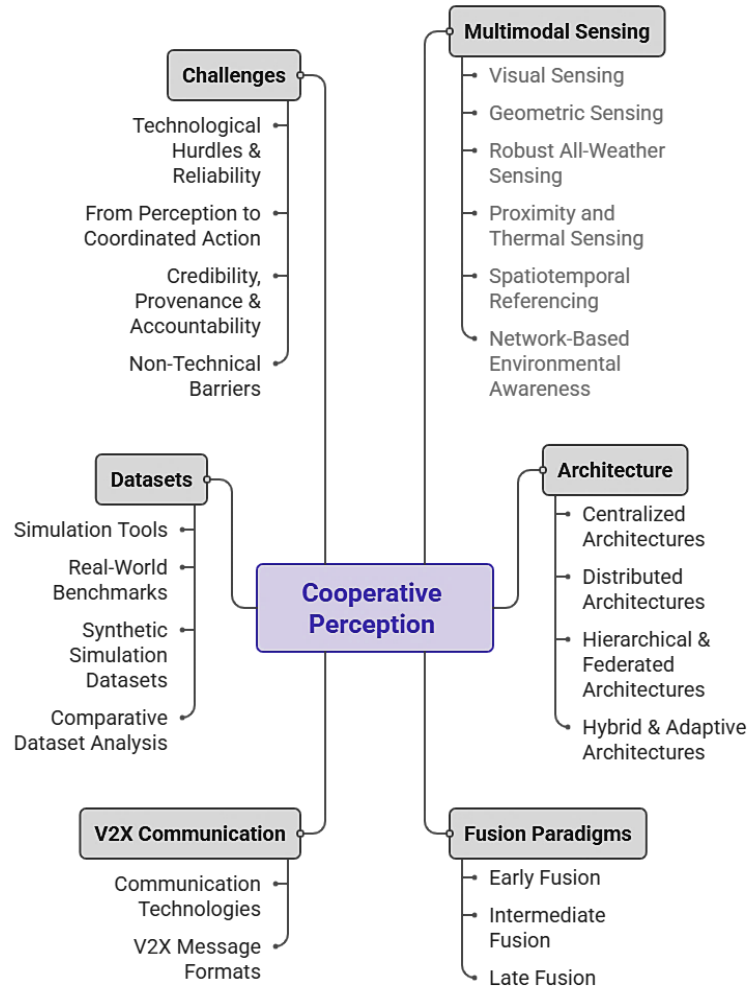


Fig. 1. Overview of the paper mindmap.

2 | Multimodal Sensing for Enhanced CPS

The capacity of automated vehicles to interpret and engage with their environment is rooted in their underlying sensor suites, which serve as the primary source of environmental data for both Advanced Driver Assistance Systems (ADAS) and high-level automation. By integrating diverse modalities—including LiDAR for spatial mapping, cameras for visual context, and radar for velocity tracking—vehicles can synthesize complementary data streams to overcome the inherent physical limitations of any single sensor. The transition from individual to cooperative perception further extends these capabilities by networking heterogeneous sensors across mobile agents, RSUs, and aerial platforms. This interconnected sensing layer establishes a distributed perceptual field that transcends traditional line-of-sight barriers, balancing critical trade-offs among precision, range, and bandwidth efficiency to provide a resilient, comprehensive model of the traffic environment.

2.1 | Visual Sensing Modalities

Cameras capture light reflected from the environment and convert it into digital images for semantic analysis [20]. High-resolution modalities, such as cameras, provide dense appearance, color, and texture details, which

are essential for object detection, lane recognition, and traffic sign identification [21]. Specialized configurations, such as fisheye cameras with ultrawide-angle lenses, achieve 360-degree coverage for comprehensive situational awareness, though they require sophisticated distortion correction for accurate data processing [22, 23]. Despite their high information density, single-agent camera systems struggle with 3D spatial estimation [24], precise velocity measurement [25], and performance degradation under adverse weather conditions [26]. A critical advancement in this field is the shift toward multi-camera cooperative networks, where asynchronous visual streams from multiple nodes are temporally aligned and fused to build unified scene graphs [27, 28]. Research suggests that such collaboration allows camera-based systems to overcome inherent occlusion and distance limitations, potentially helping camera-centric architectures overtake LiDAR in 3D detection performance [29]. The integration of Bird's-Eye-View (BEV) transformations and multi-view fusion further enhances spatial reasoning across vehicle and RSU nodes [30].

2.2 | Geometric Sensing Modalities

LiDAR generates high-resolution 3D geometry by emitting laser pulses to calculate precise distances. This process produces detailed point clouds used for mapping [31], obstacle avoidance [32], and path planning [33]. Hardware configurations vary between multi-layer spinning LiDARs, which provide dense 360-degree coverage, and solid-state LiDARs, which offer lower costs and higher refresh rates [34]. Recent developments in Frequency-Modulated Continuous-Wave (FMCW) LiDAR offer simultaneous 3D spatial data and direct velocity measurements [35]. In a cooperative context, agents utilize multi-view point cloud aggregation to compensate for sparsity and individual sensor occlusions [36, 37]. However, fusing point clouds from different agents poses unique alignment challenges due to varying poses, asynchronous timestamps, and inconsistent sampling densities [38]. These hurdles are addressed via voxel-based registration and keypoint compression before V2X transmission [39].

2.3 | Robust All-Weather Sensing

Radar sensors are indispensable for motion estimation, range measurement, and velocity estimation via Doppler shifts [40]. Their primary strength lies in maintaining stable performance under adverse weather conditions, such as rain, fog, or snow, where optical sensors may fail. While radar traditionally offers lower resolution for fine object detailing [41], modern 77 GHz and 120 GHz millimeter-wave (mmWave) radars utilize Multiple-Input Multiple-Output (MIMO) architectures to enhance angular precision [42]. Cooperative perception leverages Shared Radar Cross-Section (RCS) features and occupancy grids to estimate the motion of objects hidden from the ego-vehicle's view. Furthermore, the Integrated Sensing and Communication (ISAC) paradigm is emerging as a critical research direction, utilizing the same hardware and spectrum for both radar sensing and networking [43, 44]. This dual functionality maximizes spectrum efficiency through channel reuse and provides fine-grained situational awareness through cooperative radar imaging.

2.4 | Proximity and Thermal Sensing

Supporting sensing modalities provides critical distributed perceptual fields for specialized scenarios:

- I. Thermal and infrared cameras: These sensors visualize heat signatures, allowing for the detection of pedestrians and cyclists in low-light or night-time conditions [45]. To mitigate resolution constraints, these sensors are integrated into hybrid vision-thermal fusion frameworks within "edge-fog-cloud" architectures to maintain redundant safety coverage in multi-agent environments [46].
- II. Ultrasonic sensors: By measuring high-frequency sound wave echoes, these sensors provide near-field obstacle detection [47]. They are primarily used for low-speed maneuvering and blind-zone monitoring, providing a cost-effective solution for proximity awareness in cooperative parking and at intersections.

2.5 | Spatiotemporal Referencing

To align distributed observations into a common coordinate system, the cooperative network requires robust spatiotemporal referencing:

- I. GNSS: Global Navigation Satellite Systems provide absolute positioning [48], with accuracy reaching decimeter or centimeter levels via Real-Time Kinematic (RTK) techniques. Reliability is challenged in "urban canyons" where signal blockage occurs.
- II. IMU/INS: Inertial Measurement Units estimate motion through accelerometers and gyroscopes. They bridge GNSS outages but require periodic calibration to handle bias-induced drift [49].
- III. Cooperative Fusion and Fallbacks: Cooperative perception typically fuses GNSS/IMU data with LiDAR-based SLAM to align moving agents with static RSUs [50, 51]. To ensure continuity in GNSS-denied environments, hybrid approaches integrate Ultra-Wideband (UWB) or Dedicated Short-Range Communications (DSRC) ranging as fallback positioning mechanisms.

2.6 | Network-Based Environmental Awareness

Radio Frequency (RF)-based sensing leverages C-V2X and 5G/6G communication signals for localization and non-line-of-sight (NLOS) detection. This communication-aware sensing paradigm is vital for identifying objects obscured from traditional geometric or visual modalities. Techniques include:

- I. Received Signal Strength Indication (RSSI): Distance estimation via path-loss modeling.
- II. Time of Arrival (TOA/TDOA): Precision ranging based on signal propagation time.
- III. Angle of Arrival (AOA): Directional inference using antenna arrays for triangulation.
- IV. Fingerprinting: Matching channel characteristics against pre-built radio maps. Under the ISAC paradigm, 5G NR and 6G THz systems foster convergence in which high-frequency RF signals provide both connectivity and environmental awareness [52]. These shared RF features enable trajectory inference for NLOS objects, significantly enhancing the safety and spectrum efficiency of the cooperative network.

3 | Architecture of CPS

The structural organization of cooperative perception systems is fundamentally characterized by the intricate interconnections among sensing, fusion, and communication modules that synthesize collective environmental awareness [53]. Cooperative frameworks are defined by their specific architectural topologies, which serve as a fundamental prerequisite for mitigating spatial-temporal consistency issues and overcoming critical operational challenges, such as urban occlusions [54]. As depicted in Fig. 2, these frameworks establish a continuum of organizational paradigms—classified as Centralized, Distributed, Hierarchical, and Hybrid—that form the structural basis for resilient ITS.

3.1 | Cloud and Infrastructure-Centric Paradigms

Centralized architectures facilitate a conceptual transition from "shared sensing" to "shared perception" by structuring multi-agent evidence at a central hub to generate regionally consistent world models [54]. This paradigm is particularly effective in environments with high vehicle density and frequent occlusions, such as complex intersections or high-traffic corridors [55]. In the most established centralized model, raw data or extracted features are uplinked to remote servers to leverage massive computational resources for computationally intensive tasks [56]. These tasks include large-scale data fusion, sophisticated 3D mapping, and deep-learning-based object detection [41]. Specifically, this architecture supports long-term system learning and the generation of high-accuracy, High-Definition (HD) maps from processed LiDAR data [57]. However, the centralized paradigm is heavily reliant on high-bandwidth, stable connectivity. The inherent transmission latency remains a primary architectural constraint, rendering the cloud model potentially

unsuitable for time-critical safety applications, such as emergency braking or collision avoidance, particularly in low-connectivity rural environments [58].

The "edge-assisted centralization" model refines this approach by shifting the computational burden from distant cloud servers to localized RSUs or Multi-access Edge Computing (MEC) servers [53]. In this configuration, RSUs function as localized perceptual hubs that merge heterogeneous viewpoints into a unified world coordinate system [53]. This proximity to the data source significantly reduces both latency and bandwidth costs, which are essential requirements for real-time operation in high-speed traffic environments [59, 60]. These infrastructure units generate calibrated, time-stamped world models that are disseminated to vehicles as safety-critical cues, bridging the gap between raw sensing and downstream decision-making [61]. Despite these benefits, deployment is constrained by substantial infrastructure investment and maintenance costs. Furthermore, extreme vehicle densities in urban environments can potentially overwhelm the processing throughput of individual edge nodes.

3.2 | Distributed and Decentralized Architectures

Decentralized architectures organize cooperation through direct V2V broadcasts, thereby eliminating the requirement for a central coordinator or fixed infrastructure hub [53]. In this paradigm, agents maintain local tracks while exchanging compact information to enlarge their overall situational awareness. Individual vehicles utilize high-performance onboard processors and specialized hardware to manage independent sensor fusion and trajectory planning. This self-reliance enhances system resilience during network interruptions or in areas with unreliable infrastructure coverage, enabling vehicles to maintain robust operational awareness without external dependence [41]. However, this model must address systemic disparities arising from heterogeneous sensor quality and varying computational capabilities across a fleet, which can lead to significant variations in perception accuracy [62].

A critical scientific value-add in these systems is the implementation of consensus-based distributed fusion algorithms [63]. These algorithms unify local sensor fusion with data from connected nodes, allowing a fleet to collectively improve the detection and tracking of non-connected targets. While this distributed approach reduces core network congestion, its performance is highly sensitive to participation rates and ad-hoc communication losses [64]. Maintaining data consistency and accurate track association across diverse ad-hoc nodes remains a complex synchronization challenge, particularly given the variable network conditions found in mobile environments.

3.3 | Hierarchical, Federated, and Adaptive Organizations

Hierarchical organizations bridge localized sensing with global situational understanding by distributing sensing, fusion, and learning tasks across a three-tier system. In this structure, Tier 1 comprises vehicles that focus on low-latency sensing and preliminary feature-level fusion for immediate obstacle avoidance. Tier 2 involves RSUs and edge servers responsible for regional data aggregation, time-critical processing, and localized track management. Finally, Tier 3 utilizes cloud servers to handle long-horizon reasoning, cross-region optimization, and global high-level strategy [8]. To balance global insights with local requirements, these systems frequently employ federated learning to distribute training tasks across tiers while maintaining data locality. This ensures the system benefits from global learning without compromising data privacy or overwhelming the communication backbone.

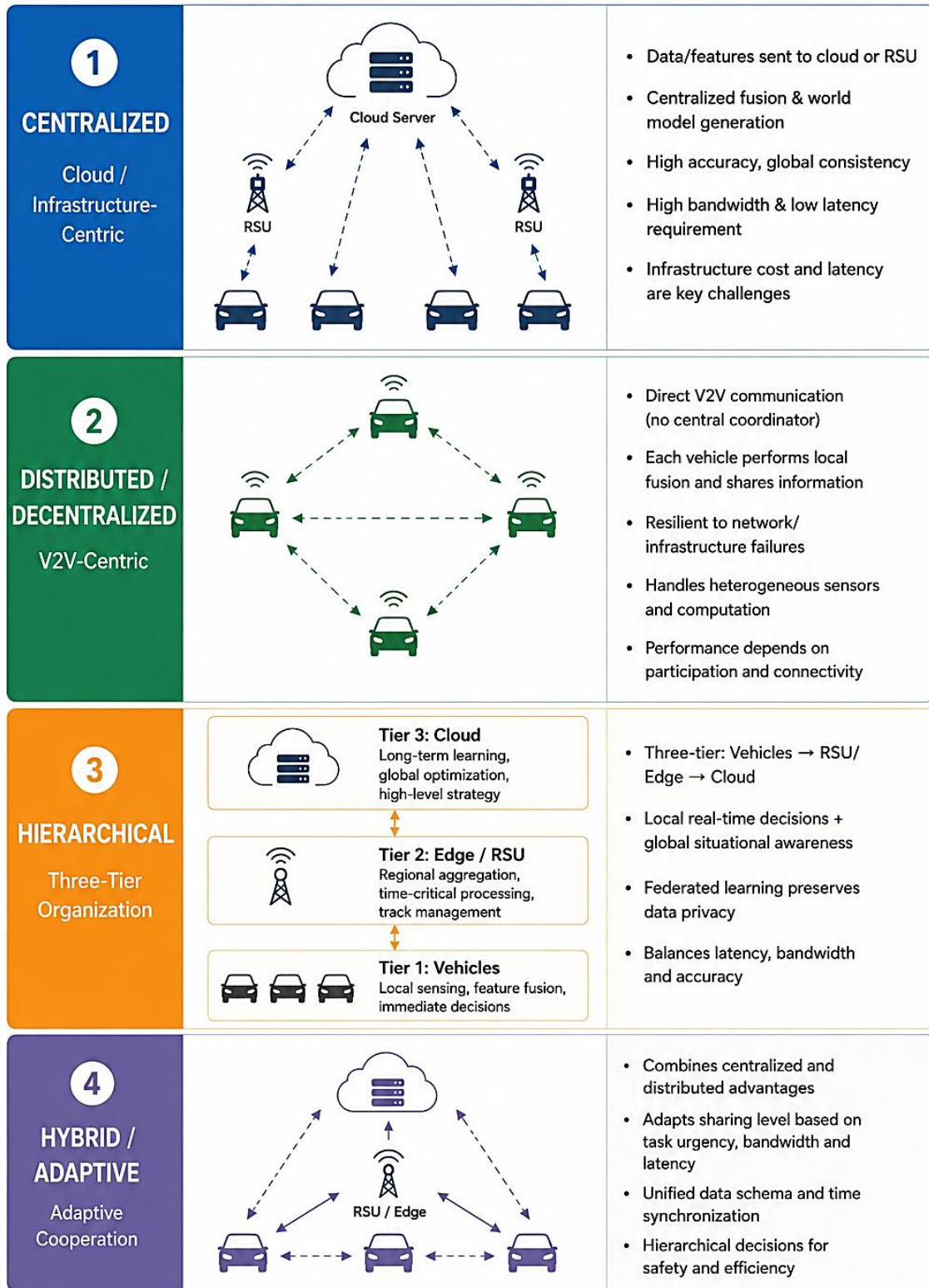


Fig. 2. Structural organization for cooperative perception systems.

Table 1. Comparative analysis of architectural paradigms for cooperative perception based on key performance metrics.

Paradigm	Latency	Scalability	Computational Requirements	Robustness
Centralized (Cloud)	~60 ms typical; up to 200 ms tail latency (field-measured 5G-to-cloud) [65]	High; integrates global/historical data, but requires ~100 Mbps uplink per vehicle [66]	Very High; data-center-scale GPU clusters [65]	Low; degrades sharply under limited cellular bandwidth [66]
Centralized (Edge/RSU)	Low; ~100 ms worst-case in field deployment [67]	Moderate; constrained by RSU density	High; MEC-grade servers, supporting sub-10 ms local inference [68]	Moderate; dependent on localized infrastructure
Distributed (Vehicle)	Very Low; processed onboard	Demonstrated with 2-5 agents in standard benchmarks; average 8.4 agents in large-scale V2X datasets [69]	Variable; onboard automotive SoC, e.g., 254 TOPS at 60 W (NVIDIA DRIVE Orin) [70]	High; 1-16% AP degradation under lossy V2V channels across representative methods [5], with ~70% performance retained at 70% packet-drop rates in some systems [71]
Hierarchical/Federated	Optimized; tiered from <10 ms (vehicle) to ~100 ms (edge) depending on task [67]	High; demonstrated scaling to 15-20 CAVs with near-constant bandwidth beyond 15 agents [72]	Distributed; balanced across multiple nodes	High; benefits from multi-tier redundancy
Hybrid/Adaptive	Adaptive; varies by task urgency	High; graceful scaling demonstrated up to 8 agents in adaptive swarm studies [73]	Balanced; leverages server and edge power	Very High; 95.3% to 76.9% task success across 0-40% packet loss in adaptive/event-triggered designs [73]

Hybrid architectures integrate centralized and distributed features to optimize performance based on task urgency, bandwidth availability, and latency [2]. In these adaptive models, communication is treated as a controllable resource within the perception loop, where the system dynamically adjusts the level of information sharing based on current network constraints and environmental complexity. Key engineering requirements for such adaptive systems include unified data schemas to align formats and timestamps across heterogeneous nodes. Furthermore, these designs implement adaptive synchronization strategies, such as data buffering, to manage network asynchrony and maintain consistency between real-time local decisions and global cloud insights. The hybrid systems balance the need for immediate local context with the broader global awareness provided by centralized nodes to enhance the autonomous system's overall safety intelligence.

4 | Fusion Paradigms

CPS are technically categorized into three distinct fusion strategies based on the specific stage at which information exchange and collaboration occur: early, intermediate, and late fusion (see Fig.3). These strategies represent a systematic progression from data-level integration to feature-level processing and, ultimately, decision-level collaboration. This taxonomy reflects the progressive integration of information as it moves through the perception pipeline, transitioning from the transmission of raw sensory observations to the reconciliation of high-level independent outputs [74]. By enabling agents to collaborate across these various

stages, CP frameworks facilitate a shared, robust environmental understanding that far exceeds the capabilities of any standalone system [75].

4.1 | Raw Data Alignment and High-Fidelity Integration

Early fusion, or data-level fusion, involves the transmission and aggregation of raw sensor data directly at the input stage. In this paradigm, agents share unprocessed or minimally processed information, such as LiDAR point clouds and camera images, to construct a holistic view of the environment. Classical systems, such as Cooper and F-Cooper, aggregate raw observations from multiple vehicles to effectively mitigate occlusions and significantly expand the collective sensing range [76]. However, the technical implementation of early fusion requires rigorous spatial alignment, precise sensor calibration, and strict temporal synchronization across all participating nodes to ensure geometric consistency across the merged multi-view reconstruction [77]. While this strategy offers the highest information fidelity by preserving raw data for joint processing, it faces substantial practical challenges, particularly regarding the high communication bandwidth required to transmit voluminous datasets. Systems employing this paradigm are also characterized by high sensitivity to network delays and pose misalignment [78]. To mitigate these communication demands, contemporary research has explored voxelized compression and semantic clustering to reduce data redundancy while maintaining the integrity of the integrated sensing environment.

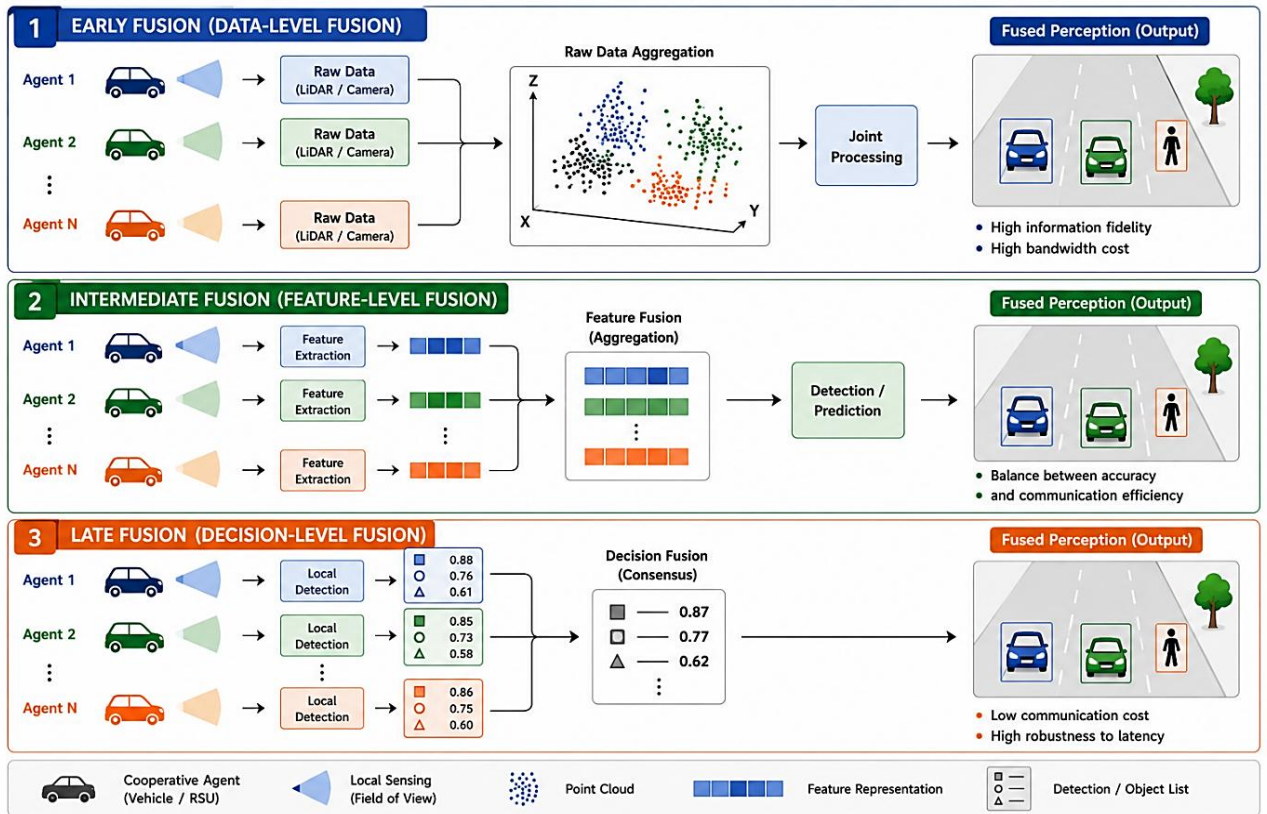


Fig. 3. Three fusion strategies based on the stage of information exchange.

4.2 | Feature-Level Cooperation and Semantic Sharing

Intermediate fusion has established itself as the mainstream approach in contemporary cooperative perception research because it addresses the critical trade-off between perception accuracy and communication efficiency. Its status as the de facto standard is supported by its dominance in academic benchmarks and real-world datasets, including OPV2V, V2X-Sim, and DAIR-V2X. In this paradigm, agents utilize neural networks to extract latent features that represent high-level semantic information. These feature representations are significantly more compact than raw sensor data—often reducing data volume by orders of magnitude—while preserving the essential spatial and semantic patterns required for complex reasoning

[79, 80]. By sharing these compact representations rather than raw pixels or points, intermediate fusion effectively reduces network congestion and bandwidth constraints, making large-scale multi-node cooperation viable for real-time applications [81, 82].

4.2.1 | Traditional intermediate fusion

Traditional methods of feature-level integration utilize basic aggregation techniques to combine information from multiple agents. These approaches rely on straightforward mathematical and convolutional operations, such as summation, concatenation, convolution, or weighted averaging, to merge locally extracted features. Representative frameworks like CoCa3D and CoBEVFusion demonstrate that these techniques can consolidate multi-agent observations into a unified representation for tasks such as 3D object detection and BEV semantic segmentation. While these operations simplify feature sharing via max or average pooling, they typically rely on spatial alignment priors and can be susceptible to feature redundancy across overlapping sensing regions [83].

4.2.2 | Attention-based intermediate fusion

The evolution of intermediate fusion has increasingly embraced Transformer-based architectures and attention mechanisms to manage the dynamic relevance between agents. Unlike static aggregation, attention mechanisms enable selective feature aggregation by assigning dynamic weights to incoming data based on its contextual importance. This global modeling capability, exemplified by models such as V2VFormer++ and MKD-Cooper, enables the system to emphasize informative features while suppressing noisy or outdated data. These architectures are particularly effective at matching multimodal features in real time, providing the adaptability and transparency required for navigating non-ideal traffic scenarios and communication dropouts [84].

4.2.3 | Topology-based intermediate fusion

Topology-based fusion strategies treat cooperative agents as nodes within a structured interaction graph, integrating spatial relations directly into the fusion process. By utilizing graph neural networks, these systems propagate information along edges defined by semantic or physical relevance, utilizing adjacency or positional embeddings to capture non-local correlations. Representative works such as V2VNet and CollabGAT use these graph-based designs to enhance both the interpretability and scalability of the collective sensing framework. This approach models the underlying communication topology to reduce dependency on a centralized synchronization authority while capturing complex interactions across heterogeneous agents [8].

4.3 | Decision-Level Aggregation and Consensus Reasoning

Late fusion, or decision-level fusion, operates by exchanging high-level processed results such as object lists, bounding boxes, or specific detection attributes. Because the shared information consists of finalized inference outputs rather than dense features or raw data, this paradigm incurs the lowest communication costs. It exhibits the highest resilience to network asynchrony and bandwidth degradation [85]. The integration process typically involves post-hoc consensus reasoning and probabilistic methods, such as confidence reweighting and object association, to reconcile overlapping predictions. Representative frameworks, such as DiscoNet, assign confidence levels to different modules and combine their outputs for final detection. While late fusion is highly efficient for resource-constrained systems, it suffers from a fundamental limitation: the loss of fine-grained scene details that cannot be recovered after local inference. Consequently, contemporary literature often benchmarks these decision-level results against high-performance feature-level architectures, such as V2VFormer++ [86], V2VNet [87], and CollabGAT [88], to assess the performance gap between simple consensus and deep feature-level cooperation.

5 | V2X Communication

5.1 | Communication Technologies

CP facilitates the deterministic exchange of environmental data among vehicles, infrastructure, and vulnerable road users, necessitating a robust V2X communication framework. The primary vehicular communication standards analyzed in the current literature are DSRC and C-V2X [89]. DSRC was specifically engineered for vehicular environments, adapting Wi-Fi technology into the IEEE 802.11p protocol to enable fast, two-way communication [89]. Operating in the 5.9 GHz band, DSRC utilizes Orthogonal Frequency Division Multiplexing (OFDM) and Carrier Sense Medium Access with Collision Avoidance (CSMA/CA) to support decentralized data exchange [14]. This architecture allows for direct communication without reliance on a central network, providing low latency for localized safety-critical applications. However, DSRC is characterized by significant technical constraints, including a communication range of under 1km and a link capacity of below 10 Mbps. Furthermore, DSRC utilizes GNSS for time synchronization via looser timing and asynchronous communication protocols.

Due to these limitations and shifting industry roadmaps, DSRC is undergoing a regulatory phase-out; for instance, the U.S. Federal Communications Commission (FCC) reallocated the 5.9 GHz spectrum in 2024, designating C-V2X as the primary technology for ITS. In contrast, C-V2X is standardized by the 3rd Generation Partnership Project (3GPP) and provides a highly scalable framework encompassing V2V, V2I, V2N, and V2P interactions. Defined in 3GPP Releases 14 through 16, C-V2X employs a dual-interface architecture comprising the PC5 and Uu interfaces [90]. The PC5 interface, or sidelink, enables direct communication in the 5.9 GHz band for safety-critical, low-latency alerts, such as crash-avoidance alerts. In contrast, the Uu interface connects vehicles to broader networks for long-range services like traffic management and infotainment. Unlike DSRC, C-V2X adheres to stricter GNSS-based time standards to ensure synchronized data fusion [91].

Despite these advancements, current implementations of both DSRC and C-V2X often fail to satisfy the rigorous requirements of CP. While they support end-to-end latencies of approximately 100ms under moderate vehicular density, CP demands more stringent quality-of-service standards to avoid temporal mismatches that degrade perceptual accuracy. The transition to 5G New Radio (NR) and the integration of mmWave spectrum and Massive MIMO are identified as critical path technologies to circumvent these existing reliability and latency bottlenecks [92].

5.2 | V2X Message Format

The functional efficacy of the V2X stack depends on standardized message formats that enable cooperative awareness and shared perception. The four critical message types utilized in these systems are Cooperative Awareness Messages (CAMs), Decentralized Environmental Notification Messages (DENMs), MAP messages, and Collective Perception Messages (CPMs) [14]. CAMs serve as a periodic "heartbeat" for each ITS station, typically broadcast at a frequency of 1-10 Hz. These messages are adaptively triggered, with transmission frequency varying based on changes in vehicle dynamics such as position, speed, and acceleration. By containing essential status data, CAMs facilitate localized safety applications, including vehicle platooning, lane change assistance, and collision avoidance [89]. While CAMs provide baseline dynamic awareness, DENMs are event-triggered messages generated only upon the detection of specific hazards, such as accidents or adverse weather [93]. DENMs are disseminated and relayed via multi-hop mechanisms by receiving stations to inform the relevant geographic area of the hazard until the event duration expires or a cancellation message is received [94].

Providing the necessary static context for these dynamic interactions, MAP messages define the road topology for intersections and road segments at a granular level [95]. These messages utilize node, link, and lane elements to represent the road geometry, enabling vehicles to orient themselves precisely within the infrastructure. When used in conjunction with Signal Phase and Timing (SPaT) data, MAP messages provide

the prerequisite structural context for applications such as Green Light Optimal Speed Advisory (GLOSA) and localization enhancement [96].

Finally, CPMs represent the perceptual extension of the V2X protocol, moving beyond the ego-vehicle status to share data regarding perceived objects and road topology. Like CAMs, CPMs are broadcasted periodically at 1-10 Hz with adaptive triggering based on the dynamics of detected objects [97]. These messages are highly structured and include mandatory containers—specifically the sensor information container and the perceived object container—as well as optional containers for originating station details. By fusing sensor-level or object-level data from multiple sources, CPMs enable a shared environmental understanding that significantly enhances the perception range and accuracy of connected automated vehicles.

5.3 | Communication Reliability and Fusion Strategy Selection

The choice of fusion strategy directly shapes how a system responds to the communication impairments discussed above. Studies of V2V channels under fading and path loss show that intermediate fusion can remain more accurate and stable under moderate channel degradation than early or late fusion, although its performance still drops sharply once the signal-to-noise ratio falls below a critical threshold [98]. Late fusion, which transmits only compact detection results rather than raw data or full feature maps, has been shown to retain roughly 70% of its clean-channel performance even at a 70% packet-dropout ratio, since losing an individual detection degrades the shared scene only partially rather than invalidating an entire feature map [N10]. Representative intermediate-fusion methods such as V2VNet and CoBEVT show average precision reductions of roughly 1-16% under lossy V2X channels, depending on the dataset and channel model [N8]. Together, these findings suggest that robustness to packet loss depends less on which fusion stage is used and more on how gracefully a given representation degrades when individual messages are lost, corrupted, or delayed. Adaptive transmission schemes that adjust the granularity or rate of shared information based on real-time channel conditions offer one practical way to mitigate these effects across fusion strategies.

6 | Datasets

While individual onboard sensors are frequently compromised by line-of-sight occlusions, restricted Fields Of View (FOV), and the degradation of point cloud density at extended ranges, CP integrates multi-agent and infrastructure-based data to achieve superior environmental awareness. The advancement of CP algorithms requires high-quality datasets that can train models to fuse multimodal information from diverse viewpoints. Due to the prohibitive costs and safety implications of real-world multi-agent field testing, research in this domain relies on a synergistic combination of high-fidelity simulation platforms and comprehensive open-source benchmarks to validate complex multi-agent interactions [41, 99].

6.1 | Simulation Tools

Simulation frameworks are indispensable for generating the multi-modal sensory data and 3D ground-truth annotations required for multi-agent CP research. The Simulation of Urban Mobility (SUMO) is a widely utilized microscopic traffic simulator known for its capacity to model large-scale, realistic traffic flows [100]. Through its Traffic Control Interface (TraCI) API, researchers can customize dynamic vehicle behaviors and integrate naturalistic trajectory data. However, SUMO remains limited in terms of sensor fidelity, as it cannot provide high-resolution, high-fidelity raw sensor data [101].

In contrast, CARLA provides a high-fidelity environment for autonomous driving research, utilizing a scalable server-multi-client architecture to distribute computational loads [102]. It offers versatility in sensor suite configurations and diverse environmental modeling, including varied weather effects and urban/rural landscapes. Despite its high rendering quality, CARLA's internal traffic manager often utilizes simplified behavioral rules that may not fully represent real-world driver interactions. To address the lack of realistic perception noise in such simulators, researchers have developed the Cyber Mobility Mirror (CMM) co-simulation platform. CMM enhances Cooperative Driving Automation (CDA) validation by generating

realistic perception data through a deep learning-based pipeline that reconstructs simulated traffic objects in 3D, thereby accounting for the "imperfect perception" often ignored in standard simulation environments [103].

OpenCDA serves as a specialized co-simulation framework that bridges CARLA and SUMO, providing a full vehicle software pipeline for CDA [104]. A significant advantage for researchers is OpenCDA's modularity, which allows substituting default perception, planning, or communication modules with custom designs to evaluate specific system components. For safety assessment, specialized tools like V2XP-ASG have been developed to generate complex adversarial scenarios for LiDAR-based models [105]. This framework employs two critical components: Adversarial Collaborator Search (ACS), which constructs adversarial collaboration graphs to identify agent combinations that lead to sub-optimal outcomes, and Adversarial Perturbation Search (APS), which utilizes black-box optimization algorithms to perturb vehicle positions and increase scene difficulty.

Further addressing safety-critical research, AUTOCASIM is built upon CARLA to evaluate cooperative models in accident-prone scenarios, such as overtaking, left turns, and red-light violations [106]. It includes an integrated networking simulation that allows customization of V2V communication dynamics. To bridge the gap between virtual testing and physical deployment, OpenCDA-ROS adapts the framework for the Robot Operating System (ROS) [107]. By reorganizing cooperative functions into a more shareable structure, OpenCDA-ROS ensures algorithmic consistency when transitioning from simulation to real-world prototyping [107].

6.2 | Real-World Benchmarks

Real-world benchmarks provide the authentic noise and environmental variability required to validate CP models. These datasets are generally categorized by their collaboration mode: V2I, V2V, or dual-mode.

The DAIR-V2X-C dataset is a large-scale benchmark for V2I research, featuring 38,845 frames of synchronized LiDAR and camera data captured at intersections [108]. Its extension, V2X-Seq, is built upon the DAIR-V2X platform to support sequential perception and trajectory forecasting across 672 hours of data [109]. While single-viewpoint roadside sensors often suffer from target occlusion, the HoloVIC dataset adopts a holographic approach. It utilizes multiple intersections equipped with synchronized configurations of cameras, LiDARs, and fisheye cameras, providing over 100,000 frames for advanced roadside perception tasks [110].

The V2V4Real dataset provides multimodal V2V data captured over 410 km of highways and city streets using two CAVs equipped with single cameras and 32-layer LiDARs [111]. However, while V2V4Real offers significant scale, its reliance on a two-vehicle setup limits its comprehensiveness for multi-agent applications involving broader swarms. LUCOOP provides an alternative multi-agent setup that captures data from three CAVs over a 4 km trajectory [112]. Although it lacks the traffic complexity of other benchmarks, it features high-precision GNSS, IMU, and UWB measurements, making it a valuable asset for evaluating SLAM and sensor domain adaptation [112].

Recent benchmarks have integrated both V2V and V2I modes to provide a holistic view of the driving environment. TUMTraf-V2X focuses on challenging intersections in Germany, capturing rare events and near-miss scenarios from both vehicle and infrastructure perspectives [79]. V2X-Real distinguishes itself as the first real-world dataset specifically designed for cross-agent cooperative tasks in congested urban settings [113]. It includes 33,000 LiDAR frames and 171,000 camera images, categorized into four sub-datasets: Vehicle-Centric, Infrastructure-Centric, V2V, and I2I CP modes, thereby supporting the development of heterogeneous collaborative models [113].

6.3 | Synthetic Simulation Datasets

Synthetic datasets provide cost-effective access to safety-critical scenarios that are difficult to capture in the real world. OPV2V is a prominent benchmark for V2V 3D object detection, featuring 70 diverse scenes and over 200,000 bounding boxes [113]. V2X-Sim offers synchronized multi-agent recordings for detection, tracking, and segmentation, though it lacks integrated noise modeling [114]. To address this, V2XSet incorporates realistic communication noise into its 11,447 frames using the OpenCDA framework [115]. Notably, both OPV2V and V2XSet are limited by the absence of non-vehicle road users (VRUs), which may impact their utility for comprehensive safety validation [115]. Environmental diversity is the hallmark of newer synthetic datasets. DOLPHINS includes six autonomous scenarios with varied weather effects, including rain and fog [116]. DeepAccident provides 57,000 frames across 12 accident-prone scenarios, supporting an end-to-end accident prediction task to forecast collision details [117]. SCOPE enhances this domain as the first synthetic multimodal dataset to incorporate physically accurate weather simulations and digital twin maps of Karlsruhe and Tübingen [118]. Finally, Adver-City was developed to bridge research gaps in adverse conditions, featuring 110 scenarios spanning heavy rain, dense fog, and glare [119].

6.4 | Comparison of Real-World Benchmarks and Simulated Datasets

We compare the datasets discussed in the previous sections into two main categories and present them in three tables.

Table 2. General comparison of CP datasets.

Feature	Real-World Benchmarks	Simulated Datasets
Primary purpose	Validate cooperative perception models under realistic traffic and sensing conditions	Train and test CP algorithms in scalable and controllable virtual environments
Data collection method	Physical vehicles and roadside infrastructure equipped with real sensors	Generated using simulation frameworks such as CARLA and OpenCDA
Sensor realism	Includes authentic sensor noise, occlusion, synchronization errors, and communication uncertainty	Sensor data are synthetically rendered and may lack realistic imperfections
Environmental control	Limited control over traffic, weather, and rare events	Full control over weather, traffic density, accidents, and adversarial conditions
Scalability	Limited by deployment cost and hardware resources	Easily scalable to thousands of scenarios
Safety-critical scenario availability	Rare and difficult to safely collect	Easily generated and repeatable
Annotation quality	Expensive and sometimes imperfect manual annotation	Automatically generated precise labels
Domain gap	Closely matches deployment environments	Suffers from sim-to-real transfer challenges
Main advantage	High realism and deployment validity	Flexibility, scalability, and low-cost experimentation
Main limitation	Expensive and limited rare-event coverage	Reduced realism compared to real traffic

Table 3. Real-world benchmark datasets.

Dataset	Collaboration Mode	Key Characteristics	Main Strength	Main Limitation
DAIR-V2X-C	V2I	38,845 synchronized LiDAR-camera frames at intersections	Large-scale roadside perception benchmark	Limited temporal sequence modeling
V2X-Seq	V2I	672 hours of sequential perception and forecasting data	Strong temporal analysis support	Mainly infrastructure-focused
HoloVIC	V2I	Multi-intersection sensing using cameras, LiDAR, and fisheye cameras	Handles occlusion effectively	Infrastructure-heavy setup
V2V4Real	V2V	410 km multimodal driving dataset with two CAVs	Real highway and urban interactions	Limited to two vehicles
LUCOOP	V2V	High-precision GNSS, IMU, and UWB measurements	Useful for SLAM and localization research	Limited traffic diversity
TUMTraf-V2X	V2V + V2I	Captures rare and near-miss intersection events	Strong safety-oriented scenarios	Limited geographic diversity
V2X-Real	V2V + V2I + I2I	33k LiDAR frames and 171k images across multiple CP modes	Supports heterogeneous collaboration	High computational complexity

Table 4. Simulated Datasets.

Dataset	Collaboration Mode	Key Characteristics	Main Strength	Main Limitation
OPV2V	V2V	70 scenes with 200k+ bounding boxes	Standard benchmark for V2V object detection	No vulnerable road users
V2X-Sim	V2X	Supports detection, tracking, and segmentation	Multi-agent synchronized data	No realistic communication noise
V2XSet	V2X	Introduces realistic communication noise using OpenCDA	Improved realism in communication modeling	Missing VRU representation
DOLPHINS	V2X	Includes rain and fog weather conditions	Environmental robustness evaluation	Limited number of scenarios
DeepAccident	V2X	57,000 frames across accident-prone scenarios	Supports accident prediction research	Focused mainly on crash scenarios
SCOPE	V2X	Physically accurate weather and digital twin maps	High environmental realism	Limited adoption due to novelty
Adver-City	V2X	Heavy rain, dense fog, and glare scenarios	Adverse-weather robustness testing	Limited long-sequence continuity

7 | Current Challenges and Future Gaps

As the initial research in CP focused on extending the sensing horizon beyond line-of-sight, the current research frontier lies in bridging the gap between visibility enhancement and operational reliability. This evolution necessitates a shift from observation-centric systems, which primarily aggregate data to resolve occlusions, to coordinated action-centric systems that ensure shared perceptual evidence translates into verifiable, safe multi-agent maneuvers.

7.1 | Technological Hurdles and System Reliability

7.1.1 | Communication constraints and environmental dynamics

The stochastic nature of vehicular communication channels represents a fundamental bottleneck for CP deployment. Hardware limitations in bandwidth and computational capacity strictly constrain the volume and frequency of high-fidelity data exchange [41]. These physical constraints are compounded by urban multi-

path effects, signal attenuation, and environmental geometry, such as tall buildings and tunnels [120, 121]. Furthermore, adverse weather conditions directly impact wireless signal propagation; atmospheric moisture and rain induce signal scattering, while fog increases path loss in radio frequency transmissions. Existing solutions often overlook the dynamic nature of this deterioration in communication. Consequently, systems are frequently forced into "selective transmission," in which discarding non-critical data can lead to suboptimal decisions and a marked reduction in perceptual accuracy.

7.1.2 | Adaptive fusion and handling heterogeneity

Traditional fusion pipelines exhibit limited efficacy in reconciling dynamic environmental variables due to their reliance on rigid data formats and fixed synchronization schemes. These frameworks fail to formalize semantic consistency when sensors encounter temporary dropouts or produce high-variance, noisy outputs [122, 123]. There is a pressing need for adaptive fusion techniques that dynamically reweight sensor contributions based on real-time quality assessments or contextual cues. Moreover, the inherent variability of sensor configurations across heterogeneous nodes presents a significant scaling challenge. Current frameworks struggle to integrate high-resolution data from premium sensor suites with the lower-resolution outputs of legacy or cost-effective hardware, necessitating new approaches to resolve multi-scale sensor heterogeneity.

7.1.3 | Expanding sensor modalities and emerging technologies

The CP research remains disproportionately concentrated on Camera and LiDAR modalities, leaving a significant gap in multi-weather robustness. Radar sensors, which maintain performance under extreme precipitation, and thermal cameras, which excel in zero-illumination scenarios, remain critically underutilized. Furthermore, the integration of emerging AI frameworks represents a vital research frontier. High-Definition (HD) maps, Large Language Models (LLMs) for processing unstructured semantic data, and end-to-end learning architectures offer untapped potential to streamline the pipeline from perception to distributed decision-making [8, 124].

7.1.4 | Security, ethics, and societal acceptance

The integrity of shared data is a prerequisite for safety. Current security frameworks, such as ETSI ITS-Secure, are often insufficient against advanced threats, such as adversarial data injection or spoofing security. Furthermore, the sharing of high-resolution sensor feeds (Camera/LiDAR) raises significant ethical dilemmas regarding pedestrian privacy and license plate identity. Although GDPR provides a general regulatory framework, it does not fully address the nuances of real-time, multi-agent sensor sharing. Finally, lagging regulatory policies regarding multivehicle interaction and decision-making liability continue to foster developer uncertainty and hinder broad societal acceptance.

Several technical approaches are emerging to address these privacy and security concerns without abandoning the benefits of cooperative perception. Federated learning allows vehicles to jointly improve a shared perception model by exchanging only model updates rather than raw sensor data, although data heterogeneity across vehicles and manufacturers can slow convergence and reduce accuracy [125]. Differential privacy can be layered onto this process by injecting calibrated noise into shared updates, offering a tunable trade-off between model utility and the strength of the privacy guarantee [126]. Secure multi-party computation offers a different kind of guarantee: rather than training a shared model, it lets vehicles jointly compute a specific result — for example, whether two vehicles are on a collision course — directly from their private sensor data, without either vehicle ever revealing that raw data to the other or to a central server. This comes at the cost of additional computation and latency compared to sharing data directly [127]. None of these techniques is a complete solution on its own, and combining them with the security frameworks discussed above remains an open engineering challenge, but together they offer a concrete technical path toward reconciling cooperative perception with the privacy and security demands identified above.

7.2 | The Paradigm Shift: From Perception to Coordinated Action

7.2.1 | Temporal Constraints

In safety-critical environments, perception outputs are insufficient when decoupled from the strict temporal constraints of the vehicle's control loop. Coordinated maneuvers must be executed within precise commitment windows defined by risk thresholds such as Time-to-Accident (TTA) and Time-to-Collision (TTC). Performance Quality of Service (PQoS) serves as the critical metric for these operations, defining the maximum allowable latency, data staleness, or confidence degradation permitted before a cooperative update is deemed unreliable for triggering interventions.

7.2.2 | Intent alignment and shared intentionality

A conceptual chasm exists between "seeing together"—the alignment of geometric primitives like bounding boxes—and "acting together". While CPS may achieve consensus on the environment's state, latent behavioral strategies often remain decentralized and unaligned. Transitioning from shared observability to shared intentionality requires formal inter-agent negotiation protocols for maneuvers such as yielding or merging. Without contextualizing perception outputs within shared decision policies, agents risk making conflicting interpretations of the same fused scene.

7.2.3 | Credibility, provenance, and accountability

Operational trust in CPS cannot be established solely based on confidence or uncertainty scores. Instead, the system must maintain rigorous data provenance—a condition-based trail that identifies the producer of the evidence and the environment in which it was generated. When conflicting perceptions emerge, the architecture must support the dynamic reallocation of trust and trace accountability to the source. This is particularly problematic in decentralized topologies, which, while preserving local autonomy, currently utilize weak coordination protocols that fail to resolve action priorities during multi-agent conflicts.

8 | Conclusion

This paper has reviewed Collective Perception Systems (CPS) as a foundational technology for next-generation Intelligent Transportation Systems. In this matter, cooperating multimodal observations from vehicles, RSUs, and infrastructure nodes into a spatiotemporally aligned shared representation is what broke through it. Architecturally, hierarchical and hybrid models have emerged as the most practical approach to deploying CPS. At the same time, feature-level fusion, particularly in attention-based and graph-structured forms, has established itself as the state of the art by balancing bandwidth efficiency against detection accuracy. The ongoing migration from DSRC to C-V2X and 5G NR is necessary, but standardized message formats must mature in parallel to carry the sensor-level evidence that genuine cooperation demands. Addressing system challenges will require formal inter-agent negotiation protocols, data-provenance mechanisms for accountability under conflict, and security frameworks hardened against adversarial injection, alongside richer benchmark infrastructure and closer co-design among perception researchers, communication engineers, and regulatory bodies.

Author Contribution

All authors contributed to the preparation of this study. J. P. drafted the manuscript, and L. Sh. provided scientific guidance, supervision, and critical review throughout the research process. All authors read and approved the final manuscript.

Funding

No funding was received to assist with the preparation of this manuscript.

Conflicts of Interest

The manuscript was not submitted to more than one journal for simultaneous consideration. The submitted work is original and has not been published elsewhere in any form or language. The manuscript has not been split up into several parts. Authors adhere to discipline-specific rules for acquiring, selecting, and processing data. No data, text, or theories by others are presented as if they were the author's own. All authors have reviewed and approved the final manuscript and consent to its publication in Big Data and Computing Visions.

References

- [1] Mendoza, A. E., Wybourn, C. A., Mendoza, M. A., Cruz, M. J., Juillard, C. J., & Dicker, R. A. (2017). The worldwide approach to vision zero: implementing road safety strategies to eliminate traffic-related fatalities. *Current trauma reports*, 3(2), 104–110.
- [2] Zhang, J., Xu, Q., Li, Z., Xu, C., & Li, K. (2025). Cooperative Safety Intelligence in V2X-Enabled Transportation: A Survey. *ArXiv preprint arxiv:2512.00490*.
- [3] Grant, E., Salmon, P. M., Stevens, N. J., Goode, N., & Read, G. J. (2018). Back to the future: What do accident causation models tell us about accident prediction? *Safety science*, 104, 99–109.
- [4] Lou, S., Stancato, G., & Piga, B. E. A. (2024). Assessing in-motion urban visual perception: analyzing urban features, design qualities, and people's perception. In *Advances in representation: new ai-and xr-driven transdisciplinarity* (pp. 691–706). Springer.
- [5] Liu, S., Gao, C., Chen, Y., Peng, X., Kong, X., Wang, K., ... others. (2023). Towards vehicle-to-everything autonomous driving: A survey on collaborative perception. *ArXiv preprint arxiv:2308.16714*.
- [6] Sánchez, M. M., van der Ploeg, C., Smit, R., Elfring, J., Silvas, E., & van de Molengraft, R. (2024). *Prediction horizon requirements for automated driving: optimizing safety, comfort, and efficiency* [presentation]. 2024 IEEE Intelligent Vehicles Symposium (IV) (pp. 2575–2582).
- [7] Zhang, L., Wang, B., Zhao, Y., Yuan, Y., Zhou, T., & Li, Z. (2024). Collaborative multimodal fusion network for multiagent perception. *IEEE transactions on cybernetics*, 55(1), 486–498.
- [8] Wei, C., Wu, G., & Barth, M. J. (2025). Cooperative perception for automated driving: A survey of algorithms, applications, and future directions. *Proceedings of the IEEE*.
- [9] Liu, S., He, M., Han, W., Chen, H., & Liu, C. (2026). Distributed Control Algorithm for Multi-Agent Cooperation: Leveraging Spatial Information Perception. *International journal of robust and nonlinear control*, 36(1), 312–328.
- [10] Li, X., Hu, E., Jin, S., Zhou, B., & Liu, J. (2026). CoGMoE: Sparse and Specialized Framework for Multi-Agent Collaborative Perception via Graph Mixture-of-Experts. *Knowledge-based systems*, 115329.
- [11] Palfy, A., Kooij, J. F. P., & Gavrilu, D. M. (2022). Detecting darting out pedestrians with occlusion aware sensor fusion of radar and stereo camera. *IEEE transactions on intelligent vehicles*, 8(2), 1459–1472.
- [12] Lu, S., & Shi, W. (2024). Sensing and data acquisition techniques. In *Vehicle computing: from traditional transportation to computing on wheels* (pp. 103–133). Springer.
- [13] Li, T., Han, X., & Ma, J. (2021). Cooperative perception for estimating and predicting microscopic traffic states to manage connected and automated traffic. *IEEE transactions on intelligent transportation systems*, 23(8), 13694–13707.
- [14] Huang, T., Liu, J., Zhou, X., Nguyen, D. C., Azghadi, M. R., Xia, Y., ... Sun, S. (2025). Vehicle-to-everything cooperative perception for autonomous driving. *Proceedings of the IEEE*.
- [15] Bejarbaneh, E. Y., Du, H., & Naghdy, F. (2024). Exploring shared perception and control in cooperative vehicle-intersection systems: A review. *IEEE transactions on intelligent transportation systems*, 25(11), 15247–15272.
- [16] Zhou, Z., Xiang, H., Zheng, Z., Zhao, S. Z., Lei, M., Zhang, Y., ... others. (2025). *V2xnp: vehicle-to-everything spatio-temporal fusion for multi-agent perception and prediction* [presentation]. Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 25399–25409).
- [17] Zhou, H., Xu, W., Chen, J., & Wang, W. (2020). Evolutionary V2X technologies toward the Internet of Vehicles: Challenges and opportunities. *Proceedings of the IEEE*, 108(2), 308–323.

- [18] Li, M., Li, H., Yao, Y., Zhu, Y., Weng, H., Jin, H., & Yang, T. (2025). Real-Time Cooperative Path Planning and Collision Avoidance for Autonomous Logistics Vehicles Using Reinforcement Learning and Distributed Model Predictive Control. *Machines*, 14(1), 27.
- [19] Ji, X. M., & Bu, F. (2025). *Cooperative trajectory planning for multiple cars via multi-agent reinforcement learning* [presentation]. Journal of physics: conference series (Vol. 3097, p. 12023).
- [20] Liu, L., Lu, S., Zhong, R., Wu, B., Yao, Y., Zhang, Q., & Shi, W. (2020). Computing systems for autonomous driving: State of the art and challenges. *IEEE internet of things journal*, 8(8), 6469–6486.
- [21] Gajjar, H., Sanyal, S., & Shah, M. (2023). A comprehensive study on lane detecting autonomous car using computer vision. *Expert systems with applications*, 233, 120929.
- [22] Kumar, V. R., Eising, C., Witt, C., & Yogamani, S. K. (2023). Surround-view fisheye camera perception for automated driving: Overview, survey \& challenges. *IEEE transactions on intelligent transportation systems*, 24(4), 3638–3659.
- [23] Jakab, D., Deegan, B. M., Sharma, S., Grua, E. M., Horgan, J., Ward, E., ... Eising, C. (2024). Surround-view fisheye optics in computer vision and simulation: Survey and challenges. *IEEE transactions on intelligent transportation systems*, 25(9), 10542–10563.
- [24] Mao, J., Shi, S., Wang, X., & Li, H. (2023). 3D object detection for autonomous driving: A comprehensive survey. *International journal of computer vision*, 131(8), 1909–1963.
- [25] Ashraf, M. H., Jabeen, F., Alghamdi, H., Zia, M. S., & Almutairi, M. S. (2023). HVD-net: a hybrid vehicle detection network for vision-based vehicle tracking and speed estimation. *Journal of king saud university-computer and information sciences*, 35(8), 101657.
- [26] Wei, C., Wu, G., & Barth, M. J. (2024). *Feature corrective transfer learning: end-to-end solutions to object detection in non-ideal visual conditions* [presentation]. Proceedings of the ieeecvfr conference on computer vision and pattern recognition (pp. 23–32).
- [27] Zhong, Y., Bi, T., Wang, J., Zeng, J., Huang, Y., Jiang, T., ... Wu, S. (2022). Empowering the V2X network by integrated sensing and communications: Background, design, advances, and opportunities. *IEEE network*, 36(4), 54–60.
- [28] Fei, L., & Han, B. (2023). Multi-object multi-camera tracking based on deep learning for intelligent transportation: A review. *Sensors*, 23(8), 3852.
- [29] Hu, Y., Lu, Y., Xu, R., Xie, W., Chen, S., & Wang, Y. (2023). *Collaboration helps camera overtake lidar in 3d detection* [presentation]. Proceedings of the ieeecvfr conference on computer vision and pattern recognition (pp. 9243–9252).
- [30] Kim, Y., Kim, S., Choi, J. W., & Kum, D. (2023). *Craft: camera-radar 3d object detection with spatio-contextual fusion transformer* [presentation]. Proceedings of the aaai conference on artificial intelligence (Vol. 37, pp. 1160–1168).
- [31] Jia, S., Gong, X., Liu, F., & Ma, L. (2025). AI-powered LiDAR point cloud understanding and processing: An updated survey. *IEEE transactions on intelligent transportation systems*.
- [32] Mochurad, L., Hladun, Y., & Tkachenko, R. (2023). An obstacle-finding approach for autonomous mobile robots using 2D LiDAR data. *Big data and cognitive computing*, 7(1), 43.
- [33] Deshpande, S., Kashyap, A. K., & Patle, B. K. (2023). A review on path planning AI techniques for mobile robots. *Robotic systems and applications*, 3(1), 27–46.
- [34] Yan, Z., Sun, L., Krajn\ik, T., & Ruichek, Y. (n.d.). *Eu long-term dataset with multiple sensors for autonomous driving*. in 2020 iee [presentation]. RSJ international conference on intelligent robots and systems (iros) (pp. 10697–10704).
- [35] Bilik, I. (2022). Comparative analysis of radar and lidar technologies for automotive applications. *IEEE intelligent transportation systems magazine*, 15(1), 244–269.
- [36] Ye, C., Wang, X., Morris, A., & Ying, Z. (2024). Pedestrian crash causation analysis and active safety system calibration. *Accident analysis \& prevention*, 195, 107404.
- [37] Ye, W., Xia, Q., Wu, H., Dong, Z., Zhong, R., Wang, C., & Wen, C. (2025). Fade3d: Fast and deployable 3d object detection for autonomous driving. *IEEE transactions on intelligent transportation systems*.
- [38] Bai, Z., Nayak, S. P., Zhao, X., Wu, G., Barth, M. J., Qi, X., ... Oguchi, K. (2023). Cyber mobility mirror: A deep learning-based real-world object perception platform using roadside LiDAR. *IEEE transactions on intelligent transportation systems*, 24(9), 9476–9489.

- [39] Wan, L., Zhao, J., Wiedholz, A., Bied, M., De Lucena, M. M., Jagtap, A. D., ... Vinel, A. (2025). A Systematic Literature Review on Vehicular Collaborative Perception—A Computer Vision Perspective. *IEEE transactions on intelligent transportation systems*.
- [40] Haider, A., Pigniczki, M., Koyama, S., Köhler, M. H., Haas, L., Fink, M., ... others. (2023). A methodology to model the rain and fog effect on the performance of automotive lidar sensors. *Sensors*, 23(15), 6891.
- [41] Bai, Z., Wu, G., Barth, M. J., Liu, Y., Sisbot, E. A., Oguchi, K., & Huang, Z. (2024). A survey and framework of cooperative perception: From heterogeneous singleton to hierarchical cooperation. *IEEE transactions on intelligent transportation systems*, 25(11), 15191–15209.
- [42] Nabati, R., & Qi, H. (2021). *Centerfusion: center-based radar and camera fusion for 3d object detection* [presentation]. Proceedings of the ieee/cvf winter conference on applications of computer vision (pp. 1527–1536).
- [43] Liu, B., Liu, J., & Kato, N. (2022). Optimal beamformer design for millimeter wave dual-functional radar-communication based V2X systems. *IEEE journal on selected areas in communications*, 40(10), 2980–2993.
- [44] Ko, S.-W., Chae, H., Han, K., Lee, S., Seo, D.-W., & Huang, K. (2021). V2X-based vehicular positioning: Opportunities, challenges, and future directions. *IEEE wireless communications*, 28(2), 144–151.
- [45] Bustos, N., Mashhadi, M., Lai-Yuen, S. K., Sarkar, S., & Das, T. K. (2023). A systematic literature review on object detection using near infrared and thermal images. *Neurocomputing*, 560, 126804.
- [46] Puppala, H., Sarasua, W., Biyaguda, S., Farzinpour, F., & Chowdhury, M. (2025). Real-time Deer Detection and Warning in Connected Vehicles via Thermal Sensing and Deep Learning. *ArXiv preprint arxiv:2509.18779*.
- [47] Nesti, T., Boddana, S., & Yaman, B. (2023). *Ultra-sonic sensor based object detection for autonomous vehicles* [presentation]. Proceedings of the ieee/cvf conference on computer vision and pattern recognition (pp. 210–218).
- [48] Kumar, A., Kumar, S., Lal, P., Saikia, P., Srivastava, P. K., & Petropoulos, G. P. (2021). Introduction to GPS/GNSS technology. In *GPS and gnss technology in geosciences* (pp. 3–20). Elsevier.
- [49] Jing, H., Gao, Y., Shahbeigi, S., & Dianati, M. (2022). Integrity monitoring of GNSS/INS based positioning systems for autonomous vehicles: State-of-the-art and open challenges. *IEEE transactions on intelligent transportation systems*, 23(9), 14166–14187.
- [50] Zhuang, C., Zhao, H., Hu, S., Feng, W., & Liu, R. (2021). Cooperative positioning for V2X applications using GNSS carrier phase and UWB ranging. *IEEE communications letters*, 25(6), 1876–1880.
- [51] Li, L., Zhang, W., Wang, X., Cui, T., & Sun, C. (2024). NLOS dies twice: Challenges and solutions of V2X for cooperative perception. *IEEE open journal of intelligent transportation systems*, 5, 774–782.
- [52] Abedi, M. R., Rashidi, Z., Mokari, N., Saeedi, H., & Zorba, N. (2025). Precise HDV Positioning through Safety-Aware Integrated Sensing and Communication in a Value-of-Information-Driven 6G V2X System. *ArXiv preprint arxiv:2510.02363*.
- [53] Yusuf, S. A., Khan, A., & Souissi, R. (2024). Vehicle-to-everything (V2X) in the autonomous vehicles domain—A technical review of communication, sensor, and AI technologies for road user safety. *Transportation research interdisciplinary perspectives*, 23, 100980.
- [54] Xiao, Z., Shu, J., Jiang, H., Min, G., Chen, H., & Han, Z. (2023). Overcoming occlusions: Perception task-oriented information sharing in connected and autonomous vehicles. *IEEE network*, 37(4), 224–229.
- [55] Dong, M., Fu, Y., Li, C., Tian, M., Yu, F. R., & Cheng, N. (2025). Task offloading and resource allocation in vehicular cooperative perception with integrated sensing, communication, and computation. *IEEE transactions on intelligent transportation systems*.
- [56] Liu, C., & Ke, L. (2023). Cloud assisted Internet of things intelligent transportation system and the traffic control system in the smart city. *Journal of control and decision*, 10(2), 174–187.
- [57] Sethuraman, R., Venkataramani, K., Yamuna Devi, M. M., & Subbaraj, S. (2025). Enhancing autonomous vehicle performance with ensemble weighted support vector-based optimization in cloud. *Cluster computing*, 28(3), 192.
- [58] Behravan, K., Farzaneh, N., Jahanshahi, M., & Seno, S. A. H. (2023). A comprehensive survey on using fog computing in vehicular networks. *Vehicular communications*, 42, 100604.
- [59] Bai, Z., Wu, G., Qi, X., Liu, Y., Oguchi, K., & Barth, M. J. (2022). *Infrastructure-based object detection and tracking for cooperative driving automation: a survey* [presentation]. 2022 ieee intelligent vehicles symposium (iv) (pp. 1366–1373).

- [60] Wang, S., Wang, Y., Chen, S., Zhou, Z., Liu, X., & Li, Z. (2025). Interactive Siamese Network-Based Roadside Perception for Multi-Vehicle Tracking. *IEEE transactions on intelligent transportation systems*.
- [61] Chen, R., Gao, L., Liu, Y., Guan, Y. L., & Zhang, Y. (2024). Smart Roads: Roadside Perception, Vehicle-Road Cooperation, and Business Model. *IEEE network*, 39(3), 311–318.
- [62] Lu, S., & Shi, W. (2023). Vehicle computing: Vision and challenges. *Journal of information and intelligence*, 1(1), 23–35.
- [63] Cai, K., Qu, T., Gao, B., & Chen, H. (2023). Consensus-based distributed cooperative perception for connected and automated vehicles. *IEEE transactions on intelligent transportation systems*, 24(8), 8188–8208.
- [64] Shi, J., Zhao, J., Zhuo, L., Wang, X., Zhan, X., & Liu, H. (2025). V2V cooperative perception with adaptive communication loss for autonomous driving. *IEEE transactions on intelligent transportation systems*.
- [65] Yang, Y., Ning, M., Shu, K., Saleh, A., Hashemi, E., & Khajepour, A. (2025). Real-World Deployment of Cloud-based Autonomous Mobility Systems for Outdoor and Indoor Environments. *ArXiv preprint arxiv:2505.21676*.
- [66] Schafhalter, P., Krentsel, A., Wei, H., Gonzalez, J. E., Ratnasamy, S., Shenker, S., & Stoica, I. (2025). TURBO: Utility-Aware Bandwidth Allocation for Cloud-Augmented Autonomous Control. *ArXiv preprint arxiv:2503.20127*.
- [67] Tsukada, M., Oi, T., Kitazawa, M., & Esaki, H. (2020). Networked roadside perception units for autonomous driving. *Sensors*, 20(18), 5320.
- [68] Peng, J., Lu, J., Zhou, Y., Cui, C., Chen, Y., & Wang, Z. (2026). OmniV2X: A Generative Foundation Planner for Efficient End-to-End Cooperative Driving. *ArXiv preprint arxiv:2606.21165*.
- [69] Wang, Y. R., Chen, S., Song, Z., & Zhou, S. (2025). WHALES: a multi-agent scheduling dataset for enhanced cooperation in autonomous driving [presentation]. 2025 ieee/rsj international conference on intelligent robots and systems (iros) (pp. 20487–20493).
- [70] NVIDIA DRIVE Orin Hardware. (2026). NVIDIA Corporation.
- [71] Fan, S., Yu, H., Yang, W., Yuan, J., & Nie, Z. (2024). Quest: query stream for practical cooperative perception [presentation]. 2024 ieee international conference on robotics and automation (icra) (pp. 18436–18442).
- [72] Wang, C., Li, Z., Li, M. F., & Wen, W. (2025). JigsawComm: Joint Semantic Feature Encoding and Transmission for Communication-Efficient Cooperative Perception. *ArXiv preprint arxiv:2511.17843*.
- [73] Qiu, A., Wang, D., Fang, Z., Partani, S., & Schotten, H. D. (2026). SwarmDrive: Semantic V2V Coordination for Latency-Constrained Cooperative Autonomous Driving. *ArXiv preprint arxiv:2604.22852*.
- [74] Ma, M., Hong, D., Zhang, H., Miao, Z., Wang, W., & Zhao, G. (2026). Enhancing collaborative perception through multi-scale contextual information integration. *Accident analysis & prevention*, 227, 108367.
- [75] Qu, D., Chen, Q., Bai, T., Lu, H., Fan, H., Zhang, H., ... Yang, Q. (2024). Sicp: simultaneous individual and cooperative perception for 3d object detection in connected and automated vehicles [presentation]. 2024 ieee/rsj international conference on intelligent robots and systems (iros) (pp. 8905–8912).
- [76] Qiao, D., Zulkernine, F., & Anand, A. (2024). Cobefusion cooperative perception with lidar-camera bird's eye view fusion [presentation]. 2024 international conference on digital image computing: techniques and applications (dicta) (pp. 389–396).
- [77] Hu, S., Fang, Z., Deng, Y., Chen, X., & Fang, Y. (2025). Collaborative perception for connected and autonomous driving: Challenges, possible solutions and opportunities. *IEEE wireless communications*.
- [78] Zhang, Y., Chen, B., Qin, J., Hu, F., & Hao, J. (2024). Coopercept: Cooperative perception for 3D object detection of autonomous vehicles. *Drones*, 8(6), 228.
- [79] Zimmer, W., Wardana, G. A., Sritharan, S., Zhou, X., Song, R., & Knoll, A. C. (2024). Tumtraf v2x cooperative perception dataset [presentation]. Proceedings of the ieeecvf conference on computer vision and pattern recognition (pp. 22668–22677).
- [80] Gao, X., Zhang, X., Lu, Y., Huang, Y., Yang, L., Xiong, Y., & Liu, P. (2024). A survey of collaborative perception in intelligent vehicles at intersections. *IEEE transactions on intelligent vehicles*.
- [81] Yu, H., Yang, W., Zhong, J., Yang, Z., Fan, S., Luo, P., & Nie, Z. (2025). End-to-end autonomous driving through v2x cooperation [presentation]. Proceedings of the aaai conference on artificial intelligence (Vol. 39, pp. 9598–9606).
- [82] Lin, C., Tian, D., Duan, X., Zhou, J., Zhao, D., & Cao, D. (2024). V2VFormer: Vehicle-to-vehicle cooperative perception with spatial-channel transformer. *IEEE transactions on intelligent vehicles*, 9(2), 3384–3395.

- [83] Gao, L., Xiang, H., Tan, T., Meng, Z., Xia, X., Xiong, L., & Ma, J. (2026). CoLoca-QuA: Cooperative Relative Pose Estimation with Query-based Attention Using Neural Intermediate Features in V2X Network. *IEEE transactions on vehicular technology*.
- [84] Guo, L., Liu, L., Tang, J., Liu, B., & Cao, S. (2026). OpenCDA-MARL: A Unified Benchmarking Framework for Cooperative Autonomous Intersection Management With Multi-Agent Reinforcement Learning. *IEEE robotics and automation letters*.
- [85] Zhang, Y., Carballo, A., Yang, H., & Takeda, K. (2023). Perception and sensing for autonomous vehicles under adverse weather conditions: A survey. *ISPRS journal of photogrammetry and remote sensing*, 196, 146–177.
- [86] Yin, H., Tian, D., Lin, C., Duan, X., Zhou, J., Zhao, D., & Cao, D. (2023). V2VFormer++: Multi-modal vehicle-to-vehicle cooperative perception via global-local transformer. *IEEE transactions on intelligent transportation systems*, 25(2), 2153–2166.
- [87] Wang, T.-H., Manivasagam, S., Liang, M., Yang, B., Zeng, W., & Urtasun, R. (2020). V2vnet: vehicle-to-vehicle communication for joint perception and prediction [presentation]. European conference on computer vision (pp. 605–621).
- [88] Ahmed, A. N., Mercelis, S., & Anwar, A. (2024). Collabgat: Collaborative perception using graph attention network. *IEEE access*, 12, 142380–142393.
- [89] George, R., Clancy, J., Brophy, T., Sistu, G., O’Grady, W., Chandra, S., ... others. (2025). Infrastructure assisted autonomous driving: Research, challenges, and opportunities. *IEEE open journal of vehicular technology*.
- [90] Ali, M. A., Khan, S. A., Aldirmaz Colak, S., Kosunalp, S., & Iliev, T. (2026). Towards 6G C-V2X Networks: A Comprehensive Survey on Mobility Management, Multi-RAT Coexistence, and Machine Learning (3M) Framework for C-ITS. *Electronics*, 15(5), 1042.
- [91] Azfar, T., Huang, K., Tracy, A., Misiewicz, S., Liu, C., & Ke, R. (2025). Traffic co-simulation framework empowered by infrastructure camera sensing and reinforcement learning. *Journal of intelligent transportation systems*, 1–19.
- [92] Sufyan, A., Khan, K. B., Khashan, O. A., Mir, T., & Mir, U. (2023). From 5G to beyond 5G: A comprehensive survey of wireless network evolution, challenges, and promising technologies. *Electronics*, 12(10), 2200.
- [93] Correia, M., Almeida, J., Bartolomeu, P. C., Fonseca, J. A., & Ferreira, J. (2022). Performance assessment of collective perception service supported by the roadside infrastructure. *Electronics*, 11(3), 347.
- [94] de Gordo, J. A. I., Alkorta, I., Urbieto, I., Velez, G., & Mujika, A. (2026). Assessing the Challenges of Collective Perception via V2I Communications in High-Speed Scenarios with Open Road Testing. *IEEE transactions on vehicular technology*.
- [95] Maksimovski, D., Festag, A., & Facchi, C. (2026). Adaptive Message Generation Rules for V2X Maneuver Coordination Service. *IEEE access*, 14, 6417–6437.
- [96] Mellegård, N., & Reichenberg, F. (2020). The day 1 C-ITS application green light optimal speed advisory – A mapping study. *Transportation research procedia*, 49, 170–182.
- [97] Castellino, R. M., Pradhan, S., & Hahn, A. (2026). Track-to-Track Fusion for Cooperative Perception Using Collective Perception Messages. *Sensors*, 26(6), 2003.
- [98] Liu, C., Chen, Y., Chen, J., Payton, R., Riley, M., & Yang, S. H. (2023). Cooperative Perception with Learning-Based V2V Communications, 2023.
- [99] Wang, N., Shang, D., Gong, Y., Hu, X., Song, Z., Yang, L., ... Lu, J. (2025). Collaborative perception datasets for autonomous driving: A review. *IEEE sensors journal*.
- [100] Gonçalves, F., Silva, G. O., Santos, A., Rocha, A. M. A. C., Peixoto, H., Durães, D., & Machado, J. (2023). Urban traffic simulation using mobility patterns synthesized from real sensors. *Electronics*, 12(24), 4971.
- [101] Ghosh, S., Mellado Sanchez, M., Sue-Ob, K., Roy, D., Jones, A., Blazquez, M. A., & Sadanandom, A. (2024). Charting the evolutionary path of the SUMO modification system in plants reveals molecular hardwiring of development to stress adaptation. *The plant cell*, 36(9), 3131–3144.
- [102] Xu, Z., Wang, X., Wang, X., & Zheng, N. (2025). Safety validation for connected autonomous vehicles using large-scale testing tracks in high-fidelity simulation environment. *Accident analysis & prevention*, 215, 108011.
- [103] Bai, Z., Wu, G., Qi, X., Liu, Y., Oguchi, K., & Barth, M. J. (2022). Cyber mobility mirror for enabling cooperative driving automation in mixed traffic: A co-simulation platform. *IEEE intelligent transportation systems magazine*, 15(2), 251–265.

- [104] Xu, R., Xiang, H., Han, X., Xia, X., Meng, Z., Chen, C.-J., ... Ma, J. (2023). The opencda open-source ecosystem for cooperative driving automation research. *IEEE transactions on intelligent vehicles*, 8(4), 2698–2711.
- [105] Xiang, H., Xu, R., Xia, X., Zheng, Z., Zhou, B., & Ma, J. (2022). V2xp-asg: Generating adversarial scenes for vehicle-to-everything perception. *ArXiv preprint arxiv:2209.13679*.
- [106] Cui, J., Qiu, H., Chen, D., Stone, P., & Zhu, Y. (2022). *Coopernaut: end-to-end driving with cooperative perception for networked vehicles* [presentation]. Proceedings of the ieeecvf conference on computer vision and pattern recognition (pp. 17252–17262).
- [107] Zheng, Z., Han, X., Xia, X., Gao, L., Xiang, H., & Ma, J. (2023). OpenCDA-ROS: Enabling seamless integration of simulation and real-world cooperative driving automation. *IEEE transactions on intelligent vehicles*, 8(7), 3775–3780.
- [108] Yu, H., Luo, Y., Shu, M., Huo, Y., Yang, Z., Shi, Y., ... others. (2022). *Dair-v2x: a large-scale dataset for vehicle-infrastructure cooperative 3d object detection* [presentation]. Proceedings of the ieeecvf conference on computer vision and pattern recognition (pp. 21361–21370).
- [109] Yu, H., Yang, W., Ruan, H., Yang, Z., Tang, Y., Gao, X., ... others. (2023). *V2x-seq: a large-scale sequential dataset for vehicle-infrastructure cooperative perception and forecasting* [presentation]. Proceedings of the ieeecvf conference on computer vision and pattern recognition (pp. 5486–5495).
- [110] Ma, C., Qiao, L., Zhu, C., Liu, K., Kong, Z., Li, Q., ... Wu, W. (2024). *HoloVic: large-scale dataset and benchmark for multi-sensor holographic intersection and vehicle-infrastructure cooperative* [presentation]. Proceedings of the ieeecvf conference on computer vision and pattern recognition (pp. 22129–22138).
- [111] Xu, R., Xia, X., Li, J., Li, H., Zhang, S., Tu, Z., ... others. (2023). *V2v4real: a real-world large-scale dataset for vehicle-to-vehicle cooperative perception* [presentation]. Proceedings of the ieeecvf conference on computer vision and pattern recognition (pp. 13712–13722).
- [112] Axmann, J., Moftizadeh, R., Su, J., Tennstedt, B., Zou, Q., Yuan, Y., ... Schön, S. (2023). *LUCOOP: leibniz university cooperative perception and urban navigation dataset* [presentation]. 2023 ieeecvf intelligent vehicles symposium (iv) (pp. 1–8).
- [113] Xu, R., Xiang, H., Xia, X., Han, X., Li, J., & Ma, J. (2022). *Opv2v: an open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication* [presentation]. 2022 international conference on robotics and automation (icra) (pp. 2583–2589).
- [114] Li, Y., Ma, D., An, Z., Wang, Z., Zhong, Y., Chen, S., & Feng, C. (2022). V2X-Sim: Multi-agent collaborative perception dataset and benchmark for autonomous driving. *IEEE robotics and automation letters*, 7(4), 10914–10921.
- [115] Xu, R., Xiang, H., Tu, Z., Xia, X., Yang, M.-H., & Ma, J. (2022). *V2x-vit: vehicle-to-everything cooperative perception with vision transformer* [presentation]. European conference on computer vision (pp. 107–124).
- [116] Mao, R., Guo, J., Jia, Y., Sun, Y., Zhou, S., & Niu, Z. (2022). *Dolphins: dataset for collaborative perception enabled harmonious and interconnected self-driving* [presentation]. Proceedings of the asian conference on computer vision (pp. 4361–4377).
- [117] Wang, T., Kim, S., Wenxuan, J., Xie, E., Ge, C., Chen, J., ... Luo, P. (2024). *Deepaccident: a motion and accident prediction benchmark for v2x autonomous driving* [presentation]. Proceedings of the aaai conference on artificial intelligence (Vol. 38, pp. 5599–5606).
- [118] Gamerding, J., Teufel, S., Schulz, P., Amann, S., Kirchner, J.-P., & Bringmann, O. (2024). *Scope: a synthetic multi-modal dataset for collective perception including physical-correct weather conditions* [presentation]. 2024 ieeecvf 27th international conference on intelligent transportation systems (itsc) (pp. 2622–2628).
- [119] Karvat, M., & Givigi, S. (2025). *Adver-city: open-source multi-modal dataset for collaborative perception under adverse weather conditions* [presentation]. 2025 ieeecvf 28th international conference on intelligent transportation systems (itsc) (pp. 740–746).
- [120] Liu, S., Hua, Y., Han, Q.-L., & Xie, L. (2026). Distributed Optimization Under Information Constraints: A Survey. *IEEE transactions on industrial informatics*.
- [121] Mondal, M. S., Ramasamy, S., Humann, J. D., Dotterweich, J. M., Reddinger, J.-P. F., Childers, M. A., & Bhounsule, P. A. (2025). Cooperative multi-agent planning framework for fuel constrained UAV-UGV routing problem. *Journal of intelligent & robotic systems*, 111(1), 12.

- [122] Qian, H., Wang, M., Zhu, M., & Wang, H. (2025). A review of multi-sensor fusion in autonomous driving. *Sensors*, 25(19), 6033.
- [123] Zhang, J., Wang, Y., Hu, J., & You, H. (2025). An Adaptive Fusion Path Tracking Strategy for Autonomous Vehicles Based on Improved ACO Algorithm. *IEEE access*.
- [124] Fang, S., Liu, J., Cui, Y., Lv, C., Hang, P., & Sun, J. (2025). Towards interactive and learnable cooperative driving automation: a large language model-driven decision-making framework. *IEEE transactions on vehicular technology*, 74(8), 11894–11905.
- [125] Zhang, Z., Liu, J., Zhou, X., Huang, T., Han, Q.-L., Liu, J., & Liu, H. (2024). On the federated learning framework for cooperative perception. *IEEE robotics and automation letters*, 9(11), 9423–9430.
- [126] Tian, Y., Shi, Y., Zhang, Y., & Tian, Q. (2024). Personalized federated learning scheme for autonomous driving based on correlated differential privacy. *Sensors*, 25(1), 178.
- [127] Lee, J., Arastehfard, A., Liu, W., Ban, X., & Hong, Y. (2026). *SecureV2X: an efficient and privacy-preserving system for vehicle-to-everything (v2x) applications* [presentation]. Proceedings of the 41st acm/sigapp symposium on applied computing (pp. 1533–1542).